Introductory Econometrics

What is Econometrics?

Why study Econometrics?

What is Econometrics?

- Econometrics is based upon the development of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy.

- Another common application of Econometrics is forecasting.

Econometrics or Statistics?

- Econometrics focuses on the problems inherent in collecting and analyzing nonexperimental data.
- Nonexperimental or observational data are collected by observing the real world in a passive way, not by controlled experiments on economic agents.
- Experimental data are often collected in laboratory environments in the natural sciences.

Why study Econometrics?

- Important to be able to apply economic theory to real world data.
- An empirical analysis uses data to test a theory or to estimate a relationship.
- A formal economic model can be tested.

Why study Econometrics?

- Theory may be ambiguous as to the effect of some policy change – You can use econometrics to evaluate the program.

- There is also a need for forecasting macroeconomic variables.

Example 1: Wage determination

- Economic model:

$$wage = f(\underbrace{educ}_{+}, \underbrace{exper}_{+}, \underbrace{training}_{?})$$

- Formal economic theory or intuition says that these factors affect productivity and hence the wage.

Example 1: Wage determination

- Econometric model:

  (a) The form of the function *f* (.) must be specified.

  (b) Other unobserved variables affecting *wage* should be deal with.

- A particular econometric model is:

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 training + u$$

Types of Data – Cross Sectional

- Cross-sectional data: Each observation is a new individual, firm, …etc. with information at a point in time.
- Cross-sectional data is a random sample.
- Key feature is that the ordering of observations does not matter for econometric analysis.
- If the data is not a random sample, we have a sample-selection problem.

Types of Data – Cross Sectional

Table 1.1

A Cross-Sectional Data Set on Wages and Other Individual Characteristics

obsno	*wage*	*educ*	*exper*	*female*	*married*
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
525	11.56	16	5	0	1
526	3.50	14	5	1	0

Types of Data – Time Series

- Time series data: Observations on a variable over time - e.g. stock prices.
- The chronological ordering of observations conveys potentially important information. Ordering matters!
- Time series data can not be assumed to be independent across time.
- Most economic and other time series are related to their recent histories.

Types of Data – Time Series

- Since not a random sample, different problems to consider.
- Features like trends and seasonality will be important.
- They can be collected at different frequencies: daily, weekly, monthly, quarterly, yearly,…

Types of Data – Time Series

Table 1.3

Minimum Wage, Unemployment, and related data for Puerto Rico

obsno	*year*	*avgmin*	*avgcov*	*unemp*	*gnp*
1	1950	0.20	20.1	15.4	878.7
2	1951	0.21	20.7	16.0	925.0
3	1952	0.23	22.6	14.8	1015.9
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
37	1986	3.35	58.1	18.9	4281.6
38	1987	3.35	58.2	16.8	4496.7

Types of Data – Pooled Cross Sections

- Pooled Cross Sections: We can pool random cross sections and treat similar to a normal cross section.
- Ordering is not crucial, but keeping track of the time for each observations is very important to account for time differences.
- Pooling cross sections from different years is often an effective way of analyzing the effects of a new government policy.

Types of Data – Pooled Cross Sections

Table 1.4

Pooled Cross Sections: Two Years of Housing Prices

obsno	*year*	*hprice*	*proptax*	*sqrft*	*bdrms*	*bthrms*
1	1993	85,500	42	1600	3	2.0
2	1993	67,300	36	1440	3	2.5
3	1993	134,000	38	2000	4	2.5
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
250	1993	243,600	41	2600	4	3.0
251	1995	65,000	16	1250	2	1.0
252	1995	182,400	20	2200	4	2.0
253	1995	97,500	15	1540	3	2.0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
520	1995	57,200	16	1100	2	1.5

Types of Data – Panel Data

- Panel Data: Time series for each cross-sectional member in a data set.
- Key feature is that the same cross-sectional units are followed over a given time period.
- This distinguishes panel data (or longitudinal data) from a pooled cross-section.
- The ordering in the cross section of a panel data set does not matter, but the ordering in the time dimension matters a lot.

Types of Data – Panel Data

Table 1.5

A Two-Year Panel Data Set on City Crime Statistics

obsno	*City*	*year*	*murders*	*population*	*unem*	*police*
1	1	1986	5	350,000	8.7	440
2	1	1990	8	359,200	7.2	471
3	2	1986	2	64,300	5.4	75
3	2	1990	1	65,100	5.5	75
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
297	149	1986	10	260,700	9.6	286
298	149	1990	6	245,000	9.8	334
299	150	1986	25	543,000	4.3	520
300	150	1990	32	546,200	5.2	493

The Question of Causality

- In testing economic theory and evaluating public policy, the economist's goal is to infer that one variable has a **causal effect** on another variable.

- Note that correlation does not imply causation, so simply establishing a relationship between variables is rarely sufficient for causality.

The Question of Causality

- The notion of **ceteris paribus** plays an important role in causality. Why?

- Because hypothesis in the social sciences are ceteris paribus in nature: All other relevant factors must be fixed when studying the relationship between two variables.

The Question of Causality

- If we have truly controlled for enough other variables in our model, then the estimated ceteris paribus effect can often be considered to be causal.

- However because of the nonexperimental nature of most data collected in the social sciences, causality can be difficult to establish.

Example 2: Returns to Education

- A model of human capital investment implies that getting more education should lead to higher earnings.
- In the simplest case, this implies an equation like:

$$earnings = \beta_0 + \beta_1 education + u$$

- The estimate of β_1, is the return to education, but can it be considered causal?

Example 2: Returns to Education

- Are all the factors affecting earnings in the equation?. What about experience or ability?

- Because not all other factors affecting earnings are held fixed when studying the relationship between *earnings* and *education*, the above equation does not control for enough variables and the relation can not be considered causal.

Example 2: Returns to Education

- The error term, *u*, includes other factors affecting *earnings*, but this is not enough here, we want to control explicitly for as much as possible.

- Note also than some variables, like *ability*, may be unobservable which can be problematic.

- More on these matters on the next lesson.